Project presentation at CONVERSATIONS 2022 – the 6th International Workshop on Chatbot Research, Applications and Design, hosted by the University of Amsterdam, Netherlands, November 22-23.

# Hungry Enough for a Chatbot: Automation Opportunities for a Restaurant Recommender

Vineeth Ravi and Jessica Staddon

J.P. Morgan Chase, AI Research

**Abstract.** We explore opportunities for transitioning a restaurant recommendation service known both for high quality recommendations and engaging social interaction (the "Text Rex" service provided by The Infatuation), from purely human-powered to at least partially chatbot-managed. We find that customer interest in cuisine and restaurant vibe are associated with conversations that are shorter and likely more suitable for a chatbot. In contrast, when customers are planning for a future outing or mention the relationships between guests in their outing (e.g., in-laws, siblings), the conversations tend to be longer, more complex, and more challenging for a chatbot.

## 1 Introduction

A customer self-service knowledge base, digital assistants for customer service agents, customer-facing chatbots, and hybrid human/machine assistant models (also known as "humbots", [15, 23]) are all ways to scale customer service. Purely automated approaches like chatbots can be negatively perceived by customers (e.g., due to communication problems and expertise gaps [14]) and transitioning from a human-operated service to a chatbot risks diminishing user trust due to perceived quality of service decline and the need for customers to adjust communication style [17]. Automation also attracts regulatory concern regarding customer expectations and awareness of automation [27].

In this paper, we describe an ongoing project to identify the customer contexts that are amenable to chatbots or humbots for a particular successful restaurant recommendation service. The recommendation service we study is "Text Rex", a text message service provided by The Infatuation [1], in March 2015-2020[1]. Users of Text Rex sent text messages specifying their interests (e.g., location, size of the dining party, cuisine preferences), and customer service agents (CSAs) responded with links to reviews of recommended restaurants. The interpersonal benefits of the Text Rex service were often called out by customers, for example, one customer said, "*It's like texting your friend who actually knows everything about food*" [2]. Indeed, customers often challenged Text Rex with complicated requests leading to long discussions. To illustrate, the following is the first customer message from a conversation with 42 customer messages in total: '*Hi! Birthday dinner recommendation. Three people on a Friday, early evening with awesome cocktails, inventive desserts (or just really great ones), memorable food that isn't spicy. Mom is paying. Work in [redacted location], going to [redacted location] after so preferably [redacted location] or [redacted location] close to trains*"[2]

Text Rex was both a fruitful channel for engaging new customers with The Infatuation and expensive to run since it was purely human-powered [2]. Automation could enable relaunching a less costly version of Text Rex but comes with the user experience challenges mentioned earlier, particularly given that some Text Rex conversations are quite complex and personal in nature. For example, in the Text Rex transcripts users often express a desire to follow up after their outing (e.g., "*Will it be weird to text you back at the end of the week and let you know how it went?*"), request their conversation stay private (e.g., "*please don't tweet this*") and ask for confirmation that they were texting with a human (e.g., "*Is this really a human?*", "*Are you a robot?*"). Indeed, when describing Text Rex in May 2019, a product lead of The Infatuation said, "*people really engaged in ways that they would never with a bot*" [2].

[1] Text Rex has paused during the covid-19 pandemic.

[2] In this paper, all quotes have neighborhood-level locations in customer messages redacted. This is not required by our organization's legal review or the Text Rex terms, but we do so to provide additional privacy for the participants while communicating the findings from our ongoing research.

To guide our exploration of chatbot-amenable Text Rex customer contexts we hypothesize that shorter Text Rex conversations are more likely to be successfully managed by a chatbot. This hypothesis is compatible with the Loebner competition's demonstration that the longer a conversation lasts, the easier it is to distinguish a chatbot from a human [3]. In addition, previous research has found that shorter chatbot conversations are associated with more positive user perception of chatbots, e.g., [21]. Given this hypothesis, the problem of identifying chatbot-amenable contexts in Text Rex, reduces to the goal of understanding how customer interests vary with conversation length and identifying interests that, when present, predict shorter conversations.

To explore the relationships between interests and conversation length we heuristically model eight popular customer interests. The models rely on string-matching, and ignore much linguistic nuance, yet they achieve strong precision and recall for these interests. Using these models, we find that cuisine preferences and an interest in the vibe of the venue are associated with shorter conversations. In contrast, customers seeking recommendations for an event the next day or later and customer mentions of the relationships between the people involved (e.g., in laws, girlfriend, etc.) are associated with longer conversations that appear more difficult for a chatbot to manage.

While our focus is a specific human-operated recommendation service (Text Rex) and interaction patterns are expected to differ in an automated service (e.g., [8,17]), our findings suggest design directions for reviving the Text Rex service as a chatbot, or hybrid human/chatbot, service. In addition, given The Infatuation's strong following and the often intimate nature of Text Rex conversation, this ongoing project is an important use case toward understanding how to design for chatbot acceptance and the ethical use of chatbots, both of which are areas highlighted as needing additional research in [11].

Overview. This paper is organized as follows. In Section 1.1 we discuss related work. In Section 2 we give an overview of the Text Rex service and the data we analyze. Section 3 describes how we define and detect customer interests. Section 4 describes our results and we conclude in Section 5 with a discussion of open problems and hypotheses.

### 1.1 Related Work

The technical challenges of automating customer service (e.g., via chatbots [18]) are well-established. Existing research develops chatbots for specific domains, including the restaurant recommendation context, e.g., [24] and studies chatbots in retail (e.g., [7,9]). Communication style and human-qualities are also important aspects of the user experience. Previous research provides strategies for designing a chatbot's communication to be appropriately human for the use case (e.g., [13,19,25,26]).

Our work is closest to the area of trust in chatbots and designing for desirable chatbot use, both areas in the chatbot research roadmap outlined in [11]. The methodology of this first phase of our project is informed by research demonstrating that chatbots are most useful when users are focused on specific, and relatively simple, tasks, as a chatbot's success in interpreting requests is closely associated with user acceptance. In particular, [28] finds that chatbots are successful with administrative tasks when users appreciate increased efficiency (e.g., by enabling multi-tasking) through chatbot use. Similarly, [4,14] find efficiency to be key in customer service, and that tolerance for chatbot errors is greater when perceived usability is high and there is an easy path to a human operator. Other research finds similar drivers of chatbot acceptance, with some variation by user characteristics, in particular, the user's level of technology acceptance [22] and an appreciation for entertaining chatbot exchanges among younger users [12].

We build on [4,14,28] in that we focus on identifying the characteristics of simpler restaurant recommendation requests for which efficiency is important. In short, our research goal is to identify the characteristics of conversations that conclude quickly.

## 2 The Text Rex Service and Data

Text Rex CSAs made use of a dining and drinking knowledge base to produce recommendations matching interests expressed by customers. When texting with customers, CSAs used labels in the knowledge base to identify appropriate recommendations. It is believed that this part of the process is easily automated as The Infatuation was able to quickly scale the Text Rex operation by hiring contractors and granting

| Speaker | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| **Customer** | *What's a good buzzing sushi spot in West Hollywood area - sake bombs and the works* | | *Sweet. Is there anything further south. Or something less pricey* | | *Done and done!* | | *Thanks!* |
| **CSA** | | *Hi there! Roku Sunset is your spot for this 9201 Sunset Blvd, West Hollywood, CA 90069. Kura is also good! 8162 Sunset Blvd., West Hollywood, CA 90046. Let us know if you need any other!* | | *Sushi Fumi is reasonably priced and a great spot for sushi! 359 N La Cienega Blvd., Los Angeles, CA 90048* | | *Enjoy!* | |

**Table 1.** An example conversation consisting of $n = 5$ customer messages in total (column 3 was sent as 2 text messages). This conversation is typical of shorter, more chatbot-amenable use cases, in which the customer has two interests (cuisine and location) and appears to be seeking a recommendation for a meal that day.

them access to the knowledge base [2]. Customers assessed the recommendations by reading the reviews sent by CSAs and asking follow up questions as needed. Text Rex CSAs only provided recommendations, not reservations.

The original Text Rex data set consists of more than 214,000 rows of text message transcripts[3]. Each row of the data set is meant to include the entirety of a single conversation between a customer and a customer service agent (CSA), and begins with a timestamp, the string "ISSUE CREATED", and ends with a timestamp and the string "ISSUE ENDED" and some message telemetry data. In each row, messages from a customer are preceded by "Customer:" or the customer's first name, and messages from a CSA are usually preceded by their first name. The customer only sees the CSA's messages and does not see the CSA's name unless the CSA discloses it.

There are no identifiers in the rows to enable conversations with the same customer to be linked, so we do not know the precise number of customers in the data set. In addition, there are no explicit customer demographics in the data set.

Over the 5 years Text Rex operated there appear to have been more than 100 CSAs. The data set only identifies CSAs by first name and some worked anonymously as "agent", so we do not have a precise count.

In the initial data set, some rows contain a merge of multiple conversations and some rows appear to be partial conversations. We experimented with a few approaches to "cleaning" the data and settled on removing any rows for which the timestamps span more than 1 day *and* for which the number of characters in the customer messages are in the top decile. We also remove rows with at most 30 characters in total across customer messages. Based on random samples of 100 rows this improves the precision of our data set from .86 to .9.[4] Our final data set consists of 178, 840 rows; the analysis of this paper is with respect to this final data set.

Privacy and Ethics. While the primary goal of the Text Rex service is restaurant reservations, some of the conversations are personal in nature. Given this, we do not share the data set, although we include a representative conversation in Table 1 as well as some excerpts, with neighborhood-level locations and names redacted. Also, an additional high-level description of the data is available from The Infatuation [2]. Our project has undergone legal review within our organization and the data controls we are using meet or exceed the privacy policy of The Infatuation (https://www.theinfatuation.com/privacy-policy).

[3] Some customers used emojis but they are not present in the transcripts.
[4] That is, .9 rows of the sample appear to be individual and complete conversations.

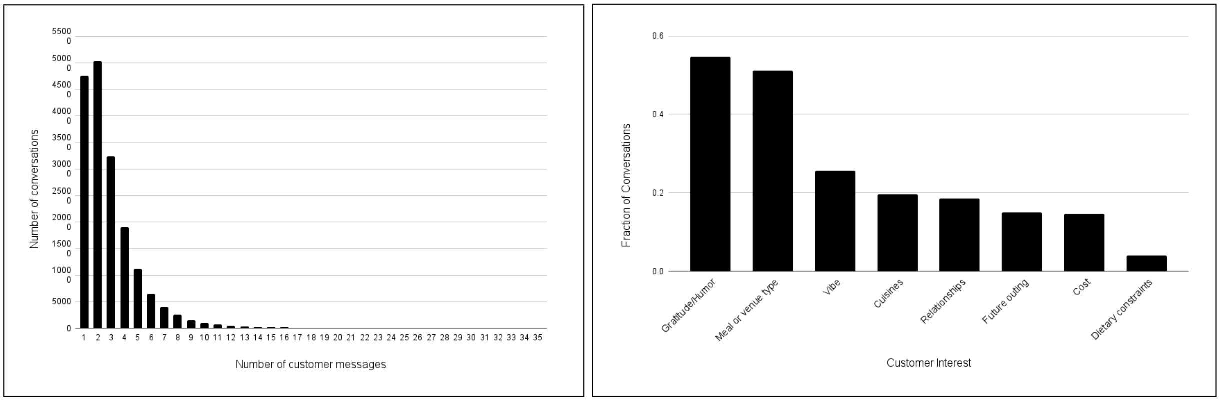


**Fig. 1.** The left figure shows the number of conversations by customer message count in the $n = 178,840$ data set. Almost 90% of conversations have 5 or fewer customer messages, but the tail is long, extending to 97 customer messages. Beyond 35 messages, the conversation counts range from 1 to 4. The right figure shows the conversation coverage of customer interests as detected by the heuristic models.

## 3 Methodology

As mentioned in Section 1, at this stage of the project we focus on identifying early indicators of conversations that may be suitable for a chatbot. We rely on conversation length as a necessary, but not necessarily sufficient, condition for chatbot-amenability since events like the Loebner competition demonstrate that extended conversation can be fatal for chatbot performance.[5] In addition, the prevalence of expressions of gratitude or humor in Text Rex conversations (Figure 1), suggests CSAs generally elicited enough customer interests to deliver appropriate recommendations.

In this section we discuss how the customers interests are defined and detected in the current phase of the project.

### 3.1 Interest definition and detection

Two labelers independently manually reviewed conversation samples and identified 8 prevalent attributes that we term customer "interests". Each interest can be expressed in many different ways (e.g., requests for Italian and Thai are both expressions of interest in cuisine), however in this stage of the project we focus on identifying whether an interest is present at all rather than its particular nature. Seven of the interests we identify are characteristics of the outing or dining event that the customer is planning, in particular: cuisine preferences, timing of the outing (i.e., is it happening soon or further in the future), meal or venue type, preferred vibe of the venue, preferred cost of the venue, dietary restrictions within the dining party, and relationships amongst the guests in the outing. The final interest is a characteristic of the conversation rather than the outing, namely, whether the conversation includes expressions of gratitude or humor. We include this interest in the analysis mostly as a sanity check since gratitude and humor are often used at the end of conversations (confirmed in Figure 2) and we are interested in the relationship between expressed interests and remaining conversation duration.

Previous research has found that customer dining interests vary in their sensitivity to context (e.g., [6]). For example, preferred price ranges may differ, or not be given at all, for milestone events like anniversaries in comparison to casual dinners. Context, may also include the customer's history, for example, a customer who has eaten several meals in loud restaurants may be more likely to request a quite venue. We note that 3 of the 7 dining-specific interests that we are consider appear less likely to be sensitive to customer's dining history: dietary interests/restrictions, relationships interest, and planning for a future outing. In Section 5 we pose hypotheses about the relationship between interests that are sensitive to history and the chatbot-suitability of conversations in which those interests are expressed.

[5] Whether the indicators we identify are *sufficient* for chatbot-amenable conversations will be the focus of future research.

We use heuristic models (i.e., regular expressions to match keywords and phrases) to detect customer interests. To define these models we iteratively evaluated them on randomly selected conversation samples of $n = 100$, and modified the keywords until achieving at least .8 precision and recall [5] for each heuristic. These evaluations treat the heuristic models as binary classifiers for the interest rather than classifiers for the particular interest values (e.g., we evaluate whether the cuisine model correctly detects that *some* cuisine was mentioned, rather than evaluating whether it correctly detects a specific cuisine, like, Italian). We took the following approach to iterative sampling:

– Precision: For each interest, a random sample of $n = 100$ is drawn from the set of conversations that the current heuristic indicates contain the interest.
– Recall: For each interest, a random sample of $n = 100$ conversation that the current heuristic model does *not* indicate contain the interest.

After reviewing the samples, precision and recall were recalculated and the heuristic models were improved by either adding keywords and phrases or refining the current keywords. Most interests required less than 4 iterations to reach at least .8 precision and recall, although some with more linguistic variation (e.g., Relationships amongst guests) required more. In general, heuristic precision is at least as high, and often higher, than recall across heuristics.

Descriptions of the heuristics follow:[6]

– Cuisines: > 30 popular cuisines (e.g., "Italian", "Thai") and menu items (e.g., "burger", "pizza")
– Future outing: > 25 phrases indicating an outing >= 1 day away (e.g., "next week", "Saturday")
– Meal or venue type: 11 meal types (e.g., "dinner", "lunch") or venue types (e.g., "bar", "drinking spot")
– Vibe: > 15 phrases describing the desired venue ambiance (e.g., "casual", "cozy", "hip")
– Cost: > 15 phrases related to the cost of menu items (e.g., "Michelin", "pricey", "cheap")
– Dietary constraints: > 10 phrases describing dietary restrictions (e.g., "vegan", "gluten free")
– Gratitude/Humor: > 10 expressions of gratitude (e.g., "thanks", "ty") and 2 expressions of humor ("haha", "lol")
– Relationships amongst guests: > 20 phrases describing relationships (e.g., "girlfriend", "dad", "colleagues")

As expected, more customer interests are detected in longer conversations. For example, the mean and median number of interests in conversations with at least 6 customer messages are 3.19 and 3, respectively, versus a mean of 1.9 and a median of 2 for conversations with at most 5 customer messages. The subsequent analysis focuses on understanding which interests are associated with shorter, more chatbot-amenable conversations, and how interests predict such conversations.

## 4 Findings

Text Rex conversations are predominantly positive, indeed the fraction of messages containing explicit customer gratitude or humor (55%) exceeds even the mentions of meal or venue type (51%). The service also appears to have a different value proposition from recommenders like Yelp, that emphasize ratings and cuisine [16], as customers are more likely to express an interest in a restaurant's vibe (26%) than cuisine (20%), which is closely followed by an interest in the relationships amongst those in an outing (18%).

Conversation length varies with customer interests. For example, while more than 55% of conversations in which a cuisine interest appears, end after just 1 additional customer message, that percentage is less than 48% in the case of conversations in which relationships interests appear or the customer mentions they are planning for a future event. Figure 2 shows that gratitude/humor is the interest that most signals the conversation will soon end (since it is customary to wrap up a conversation with gratitude) and in contrast, planning a future event and customer mentions of the relationships between guests in the outing are most associated with additional conversation.

[6] In this initial phase of the project we focus on interests that lend themselves to heuristic models, leaving more complicated interests, like location, for later phases.

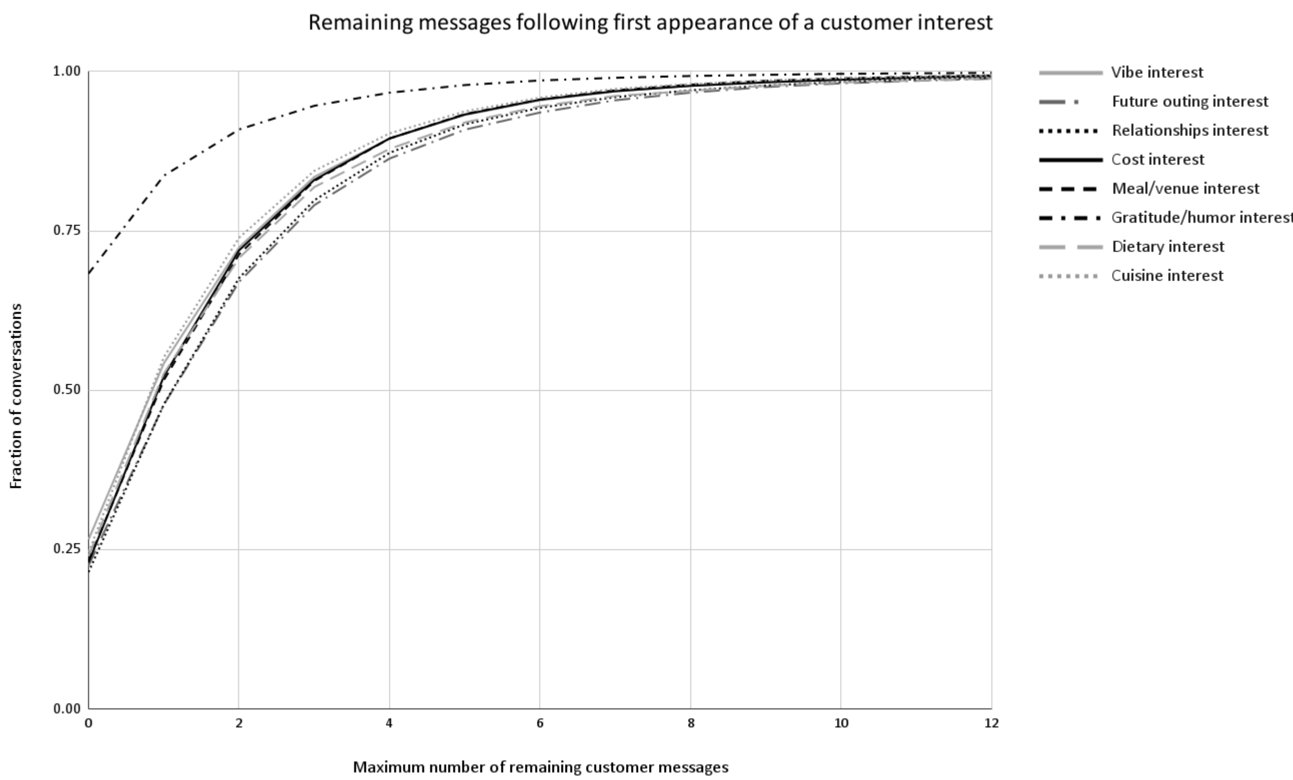


**Fig. 2.** For a given number of remaining messages, this chart shows the fraction of conversations that conclude in at most that many messages following the first appearance of an interest. Gratitude/humor is the best indicator of a conversation nearing conclusion, whereas planning for a future outing and mentions of the relationships amongst those in the outing, are more indicative of continuing conversation.

In addition, interests predictive of total conversation length may appear early in the conversation. For example, for conversations with more than 5 customer messages, the customer mentions planning for a future event within the first 2 messages in 18.8% of the conversations, but in conversations with at most 5 customer messages, a future event is only indicated in the first or second message in 12.6% of the conversations. Similarly, almost 22% of conversations with more than 5 customer messages mention relationships amongst guests within the first 2 messages, versus less than 16% of the conversations with 5 or fewer customer messages.

Note that interests that are likely less sensitive to variations in dining history (e.g., dietary restrictions, relationships, planning for future outings) are more associated with longer conversations. A potential explanation for this is that history-sensitive interests indicate a customer is planning for an imminent outing, hence otherwise they might not be able to specify those history-sensitive interests. For example, while a customer may know they are craving sushi for dinner, they are less likely to know they will be craving sushi 2 weeks from now. Imminent outings may lead to shorter conversations out of necessity. In contrast, when a customer emphasizes less history-sensitive interests (e.g., dietary restrictions, relationships between guests) they may be planning a future outing and thus be available, and perhaps interested in, a longer conversation. We highlight this hypothesized explanation as an open problem in Section 5, since understanding customer expectations can facilitate the successful transition from a human-powered to chatbot or humbot service.

Finally, to further explore how customer interests are associated with conversation length, we train an initial gradient-boosted decision tree ensemble learning model to predict the number of customer messages (97 classification categories; conversations range from 1 to 97 customer messages)[7]. The model takes the customer interests (binary classifiers) along with the message when an interest first appears, as features. Across 5-fold cross validation (80/20 split of training and test data), the model achieves more than 65% classification accuracy [5]. We note that it is likely this accuracy can be improved by optimizing the parameters and including additional information about when interests appear (beyond the first appearance) as features.

## 5 Conclusion & Open Problems

We present initial findings from this ongoing project and many open problems remain. In particular, we have relied on heuristic models that do not capture the nuances of language (e.g., negation, sarcasm) well and some of the interest categories are quite broad (e.g., all cuisines). Because of this, messages for which the same interests are detected may vary a lot in complexity. For example, a conversation in which a customer requests an Italian restaurant recommendation and one in which a customer specifically requests to not have Italian cuisine, will appear the same in terms of our heuristic models, but the latter is a much less specific request and may lead to a longer conversation. In addition, some interests like location, which manual review of the data suggests is predictive of time-sensitive requests, are not modeled well with string-matching heuristics and so are not part of our analysis. In future work, we will explore using transformer-based language models (e.g., [10, 20] ) to better capture linguistic nuance and detect a wider variety of customer interests.

We have focused on predicting conversation length since previous work suggests shorter conversations tend to be more easily supported by chatbots. We emphasize that conversation length is more likely a necessary than sufficient condition for chatbot suitability. In future work we plan to pilot a chatbot and explore what additional characteristics help predict a successful chatbot experience in the restaurant recommendation context.

Also, our data set is not perfectly clean and is certainly influenced by the high quality of service provided by the Text Rex CSAs. That said, the analysis yields design directions useful for developing and testing a chatbot for restaurant recommendations. In particular, our analysis suggests a restaurant chatbot is most likely to be successful with customers who are planning for an outing in the near future and/or have cuisine preferences. Customers with more nuanced constraints related to the relationships of guests in the outing or dietary preferences are more likely to need longer conversations that a chatbot may have difficulty managing. To provide the right form of assistance to each customer, the app or web

[7] Implemented via the scikit-learn and xgboost machine learning libraries.

UI could highlight the chatbot for those with urgent, and/or less constrained, recommendation needs, or alternatively, the data suggest it may be possible to detect that a customer can be well-served by a chatbot based on their initial messages.

Finally, while our focus is understanding the Text Rex customer contexts that may be suitable for a chatbot, our project suggests two hypotheses that we plan to explore in future work and that are relevant to other restaurant recommendation contexts:

*Hypothesis 1*: A focus on interests that are sensitive to a customer's dining history (i.e., preferences that are more likely to change from meal to meal) is indicative of a recommendation need that is easier for a chatbot to satisfy.

*Hypothesis 2*: When transitioning from a human-powered to chatbot/humbot recommendation service, customers expressing interests that are history-sensitive and meal-specific, may be more tolerant of an automated service.

The first hypothesis is motivated by our finding that customer mentions of interests like cuisine and vibe are often associated with shorter conversations and hypothesis 2 is motivated by the conjectured explanation in Section 4 that the urgency of a customer's recommendation needs may impact their acceptance of automation, i.e., that they may need to be "hungry enough for a chatbot".



## References

1. The Infatuation. https://www.theinfatuation.com.
2. Jesse Rose, The Infatuation & Zagat. https://www.youtube.com/watch?v=JfSMD9yPznU.
3. The Loebner Prize, 1990-2020. https://www.ocf.berkeley.edu/ arihuang/academic/research/loebner.html.
4. M. Ashfaq, J. Yun, S. Yu, and S. M. C. Loureiro. I, chatbot: Modeling the determinants of users' satisfaction and continuance intention of ai-powered service agents. *Telematics and Informatics*, 54:101473, 2020.
5. R. Baeza-Yates, B. Ribeiro-Neto, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.
6. B.-L. Chua, S. Karim, S. Lee, and H. Han. Customer restaurant choice: an empirical analysis of restaurant types and eating-out occasions. *International journal of environmental research and public health*, 17(17):6276, 2020.
7. R. D. Cicco, S. C. L. d. Costa e Silva, and R. Palumbo. Should a chatbot disclose itself? implications for an online conversational retailer. In *International Workshop on Chatbot Research and Design*, pages 3–15. Springer, 2020.
8. E. A. Croes and M. L. Antheunis. 36 questions to loving a chatbot: Are people willing to self-disclose to a chatbot? In *International Workshop on Chatbot Research and Design*, pages 81–95. Springer, 2020.
9. R. De Cicco, S. Iacobucci, A. Aquino, F. Romana Alparone, and R. Palumbo. Understanding users' acceptance of chatbots: An extended tam approach. In *International Workshop on Chatbot Research and Design*, pages 3–22. Springer, 2021.
10. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
11. A. Følstad, T. Araujo, E. L.-C. Law, P. B. Brandtzaeg, S. Papadopoulos, L. Reis, M. Baez, G. Laban, P. McAllister, C. Ischen, et al. Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, 103(12):2915–2942, 2021.
12. A. Følstad and P. B. Brandtzaeg. Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience*, 5(1):1–14, 2020.
13. A. Følstad, C. B. Nordheim, and C. A. Bjørkli. What makes users trust a chatbot for customer service? an exploratory interview study. In *International conference on internet science*, pages 194–208. Springer, 2018.
14. A. Følstad and M. Skjuve. Chatbots for customer service: user experience and motivation. In *Proceedings of the 1st international conference on conversational user interfaces*, pages 1–9, 2019.

15. J. Grudin and R. Jacques. Chatbots, humbots, and the quest for artificial general intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.
16. A. Hicks, S. Comp, J. Horovitz, M. Hovarter, M. Miki, and J. L. Bevan. Why people use yelp. com: An exploration of uses and gratifications. *Computers in Human Behavior*, 28(6):2274–2279, 2012.
17. J. Hill, W. R. Ford, and I. G. Farreras. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in human behavior*, 49:245–250, 2015.
18. K. Kvale, O. A. Sell, S. Hodnebrog, and A. Følstad. Improving conversations: lessons learnt from manual analysis of chatbot dialogues. In *International workshop on chatbot research and design*, pages 187–200. Springer, 2019.
19. C. Liebrecht, L. Sander, and C. v. Hooijdonk. Too informal? how a chatbot's communication style affects brand attitude and quality of interaction. In *International Workshop on Chatbot Research and Design*, pages 16–31. Springer, 2020.
20. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.
21. E. Luger and A. Sellen. " Like Having a Really Bad PA" the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 5286–5297, 2016.
22. C. B. Nordheim, A. Følstad, and C. A. Bjørkli. An initial model of trust in chatbots for customer service—findings from a questionnaire study. *Interacting with Computers*, 31(3):317–335, 2019.
23. Y. Oshrat, Y. Aumann, T. Hollander, O. Maksimov, A. Ostroumov, N. Shechtman, and S. Kraus. Efficient customer service combining human operators and virtual agents. *arXiv preprint arXiv:2209.05226*, 2022.
24. N. Sardella, C. Biancalana, A. Micarelli, and G. Sansonetti. An approach to conversational recommendation of restaurants. In *International Conference on Human-Computer Interaction*, pages 123–130. Springer, 2019.
25. M. Skjuve, I. M. Haugstveit, A. Følstad, and P. B. Brandtzaeg. Help! is my chatbot falling into the uncanny valley? an empirical study of user experience in human-chatbot interaction. *Human Technology*, 15(1), 2019.
26. N. Svenningsson and M. Faraon. Artificial intelligence in conversational agents: A study of factors related to perceived humanness in chatbots. In *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*, pages 151–161, 2019.
27. M. Veale and F. Z. Borgesius. Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4):97–112, 2021.
28. J. Zamora. I'm sorry, Dave, I'm afraid I can't do that: Chatbot perception and expectations. In *Proceedings of the 5th international conference on human agent interaction*, pages 253–260, 2017.